5-2020

Patterns of Genomic Introgression in Topminnow Hybrid Zones

Amanda F. Ray

The University of Southern Mississippi

Patterns of Genomic Introgression in Topminnow Hybrid Zones

by

Amanda Ray

A Thesis
Submitted to the Honors College of
The University of Southern Mississippi
in Partial Fulfillment
of Honors Requirements

May 2020

Approved by:

_____

Jacob F. Schaefer, Ph.D., Thesis Adviser
Professor of Biological Sciences

_____

Emily Clark, Ph.D., Thesis Adviser
Assistant Teaching Professor of Biological Sciences

_____

Jacob F. Schaefer, Ph.D., Director
School of Biological, Environmental, and Earth Sciences

_____

Ellen Weinauer, Ph.D., Dean
Honors College

Abstract

Hybridization and introgression are two important evolutionary mechanisms that can increase genetic diversity. Interesting introgression patterns can form when parental species have genes that confer some adaptive benefit to the organism. The *Fundulus notatus* species complex contains species with various identifying characterisics. *Fundulus notatus*, the blackstripe topminnow, and *Fundulus olivaceus*, the blackspotted topminnow, are closely related and occupy many of the same rivers in their preferred niches. These two species often hybridize and form hybrid zones where their niches overlap. We studied two hybrid zones located in the Tombigbee River and Spring River. Within each hybrid zone, we performed a structure analysis to identify parental species and those of mixed ancestry. We then performed a sliding window analysis to analyze the hybrid genomes present in each hybrid zone. When comparing the two hybrid zones, we identified overlapping regions of low introgression on five chromosomes, whereas we only found one overlapping region of high introgression. We then identified putative functions of genes present within these regions, and most genes in the regions of low introgression had functions in the categories of cell development, intra/intercellular transport, and cell signaling.

Keywords: hybrid zones, Fundulus, introgression, hybridization, single nucleotide polymorphism (SNP), evolution

Acknowledgements

I would like to thank my amazing advisors, Dr. Jake Schaefer and Dr. Emily Clark. Thank you for taking me under your wing(s) and encouraging me every step of the way. I have gained so much valuable knowledge about not only biology, evolution, and the topminnow, but also about collaboration and curiosity. You both took the time to answer my many questions and ensure that I was confident, and for that I will be forever grateful!

Thank you to Dave Duvernell of Missouri S&T for field collections alongside Dr. Schaefer, as well as his lab for processing the collected fin clips to prepare the GBS data.

I would also like to thank my best friends, my "keefee krewe", for keeping me sane throughout this process. Having friends going through the same process with me was so valuable, and I cannot thank you all enough for the support you have shown me.

To my family, thank you for supporting me and encouraging me even though you don't really know much about biology. I greatly valued our many phone calls about how my project was going. To Lucas, thank you for always answering the phone and being a light when I thought I was going crazy. I would not have made it this far without you.

Table of Contents

List of Tables

List of Illustrations

List of Abbreviations

SNP	Single nucleotide polymorphism
GBS	Genotyping-by-sequencing
F_{st}	Fixation index
MHC	Major histocompatibility complex

Chapter 1: Introduction

Hybridization and introgression are two relatively common events that typically go hand in hand and are important parts of the evolutionary process (Arnold, 1992; Abbott et al., 2013). Hybridization occurs when individuals of different species crossbreed, and introgression follows as genes transfer between the two parental gene pools (Twyford and Ennos, 2012). Hybrid zones are often formed between species that undergo natural secondary contact (Haines et al., 2019). Hybridization can occur naturally, but it is often found in areas of environmental degradation, which often include anthropogenic changes to habitats. Barriers to hybridization typically involve extrinsic or intrinsic isolating mechanisms, including reproductive barriers (Mallet, 2005). Reproductive isolating barriers inhibit the amount of hybridization and introgression, so they act as a filter to gene flow (Twyford and Ennos, 2012). Reproductive isolation can be influenced by intrinsic genetic incompatibilities that decrease fitness of any hybrids produced (Schaefer et al., 2016). Intrinsic isolating barriers are those that are within or of the organism, and examples include hybrid inviability or sterility. Extrinsic isolating barriers exist outside of the organism as physical barriers, an example being spatial separation, which does not allow for reproduction (Mallet, 2005). When these barriers are removed or do not exist, widespread gene flow may occur so much so that introgression is "invasion of the genome" (Mallet, 2005; Twyford and Ennos, 2012).

Introgression can occur randomly throughout the genome, or it can be variable. Regions of the parental genome that increase the fitness of the offspring are more likely to remain present in the species. For example, fragments of the Neandertal genome are

present in the DNA of modern humans (Vernot & Akey, 2014). Many of these genes are involved in the integumentary system, which suggests that these Neanderthal genes provided modern humans with adaptive variation for skin phenotypes (Vernot & Akey, 2014). Alleles that contribute to the barriers between species will have lower rates of introgression and are more likely to remain together (Larson et al., 2013). Selection pressures in a given hybrid zone can differ for phenotypes, leading to variation in gene introgression. Alleles that give the hybrid higher fitness should spread more quickly than those that are not beneficial for the hybrid (Haines et al., 2019). If few loci are under divergent selection in the two species, these loci are likely to remain distinct and together (Mallet, 2005). Unlinked or distantly linked regions that are equally fit in either genomic background should be able to flow relatively freely between species (Mallet, 2005; Larson et al., 2013). There is good evidence that hybridization and introgression increase diversification rates. These processes can produce "hopeful monsters", as hybrids can have adaptive sets of genes that make them substantially different from either parental species and give them a greater fitness (Dittrich-Reed & Fitzpatrick, 2013). Genetic analysis of hybrid zones has been a developing process for many years. Recent mechanisms to study hybridization include analysis of single nucleotide polymorphisms (SNPs) by comparing genome sequences. These genomic comparisons can identify genes and gene combinations that maintain the hybridizing species as discrete entities. It also shows which genes and combinations are introgressed and which have an adaptive significance (Abbott et al., 2016).

The *Fundulus notatus* species complex includes the blackstripe topminnow (*F. notatus*), the blackspotted topminnow (*F. olivaceus*), and the broadstripe topminnow (*F.*

euryzonus) (Duvernell et al., 2007; Schaefer et al., 2016). These are different types of killifish. Killifish are generally non-migratory and have large effective population sizes, making them good arenas for natural selection and genetic variation (Waits et al., 2016). These killifishes are all closely related and found in similar niches and ranges. *Fundulus notatus* and *F. olivaceus* have a broadly overlapping distributions, which has resulted in the formation of replicate hybrid zones among drainage basins (Duvernell et al., 2007). These two species are generally morphologically similar throughout their ranges; however, important life-history strategies distinguish the species. For instance, *F. olivaceus* attains a slightly larger adult size and produces a smaller number of larger eggs, whereas *F. notatus* invests more in reproduction (as indicated by gonadal somatic index) and less in growth, resulting in reduced body size and somatic condition (Schaefer et al., 2016). Another character that distinguishes the two species is black dorsolateral spots, which are found primarily on *F. olivaceus* and are sexually dimorphic (more prominent in males). A key cytogenetic difference between the species is their diploid number of chromosomes. *F. olivaceus* has 48 chromosomes, whereas *F. notatus* has 40 chromosomes in most of its range. In the Tombigbee River, *F. notatus* actually has 44 chromosomes (Duvernell et al., 2007; Black & Howell, 1978). In regions occupied by both species, *F. olivaceus* usually occupies high-gradient, clear, gravelly headwater streams, and *F. notatus* is typically found downstream in quieter sloughs and backwaters (Duvernell et al., 2007; Schaefer et al., 2016). Hybrid zones between *F. olivaceus* and *F. notatus* are usually formed near confluences between smaller creeks and larger rivers where the habitat shifts from one species' niche to that of the others (Schaefer et al., 2016).

We studied two hybrid zones between *F. olivaceus* and *F. notatus*, one located in the Tombigbee River (confluence of Kings Creek and Town Creek, Tupelo, Mississippi) and one in the Spring River (below Lowell Reservoir, west of Joplin, Missouri). When comparing these two hybrid zones, we focused on a few topics and questions. Are the rates of hybridization and introgression the same in the two hybrid zones, and is hybridization between two species always the same at replicate sites? Do regions of the genome introgress at different rates? Are patterns of introgression across the genome the same in the two hybrid genomes? What functional roles do we see in areas of the genome with high or low rates of introgression? We generally expected to see patterns of low introgression in regions of the genome that contain genes that function in maintaining the species' boundaries, such as cell development, cell cycle, and transcription.

Chapter 2: Methodology

We sampled two previously characterized hybrid zones between populations of *F. notatus* and *F. olivaceus* (Schaefer et al., 2011; Duvernell & Schaefer, 2013; Duvernell et al. 2013) in the Spring River (below Lowell Reservoir, west of Joplin, Missouri 37.062756 -94.708679) and Tombigbee River (confluence of Kings Creek and Town Creek, in Tupelo, Mississippi 34.234282 -88.69602). For both systems, the center of the contact zone was known from previous studies, and had been sampled intensively in an attempt to capture as many individuals as possible (110 in the Spring River and 192 in the Tombigbee) from admixed populations. Sampling was done using dipnets, and fin clips were preserved in either 100% ethanol or a high salt preservative (Seutin et al., 1991). Genomic DNA was extracted using the Qiagen DNeasy Blood and Tissue Kit (Qiagen). The EcoT22I restriction enzyme was used to construct genotyping-by-sequencing (GBS) libraries following Elshire et al. (2011) and libraries were sequenced using Illumina HiSeq. Raw GBS data was analyzed with the TASSEL (version 5.0) pipeline (Glaubitz et al., 2014) using a *Fundulus olivaceus* genome as a reference (Brennan et al., 2018), and aligning with Bowtie 2.0 (Langmead & Salzberg, 2012). After genotyping calls were made, data were filtered for missing data (maximum of 25% missing by individual and 10% missing by locus) and only biallelic loci were retained. We also removed any loci with excess artifactual heterozygosity (loci with >50%) (Schaefer et al., 2016). We performed an initial screen of all individuals to assess admixture status by conducting a STRUCTURE analysis (Pritchard et al., 2000), with $K = 2$ (average of five replicates of one million repetitions after a burnin of 100,000). In this analysis K corresponds to the number of populations, or separate gene pools. Individuals with coefficient of

membership (*Q*-score) > 0.97 or less than 0.03 were designated as parental species (*F. notatus* or *F. olivaceus*) and used in later analyses.

The *F. olivaceus* reference genome is currently organized into 1,200 well defined linkage groups or contigs and is not annotated. In order to characterize genes associated with our SNPs, we used the annotated and mapped genomes for the sister species *F. heteroclitus* (Waits et al., 2016). We used MUMmer (Marçais et al., 2018) to align the *F. olivaceus* reference (1,200 contigs) to the annotated reference genome for *F. heteroclitus.* In aligning *F. olivaceus* contigs to the *F. heteroclitus* reference, we only retained contigs with multiple alignments exclusively to one of the 24 linkage groups (chromosomes). Using these high quality alignments, we then assigned each of our SNPs a position along one of the 24 chromosomes. With SNPs arranged along chromosomes, we then used VCFtools (ver. 1.16) to identify F_{st} (fixation index) outliers using a sliding window analysis (windows 500,000 base pairs wide, in 100,000 base pair steps). The fixation index, F_{st}, provides a measure of genetic differentiation in a population.

For the Tombigbee and Spring River hybrid zones, the goal was to identify genes and their respective functions in areas of the genome exhibiting low and high introgression, respectively. For genomic regions of low introgression, the average F_{st} value was 0.91 and 0.90, respectively. Because of the overall high level of population differentiation in both zones, we confined our analysis to areas of the genome with F_{st} values at or exceeding 0.99 for the Tombigbee River, and 0.985 for the Spring River, indicating little to no introgression. For genomic regions showing high introgression, we confined our analysis to genomic regions with F_{st} values equal to or less than 0.03. Of these regions, we chose regions with 10 or more SNPs in at least 100,000 base pairs for

further analysis. This was decided so that there was more genetic variation present in each region than lower numbers of SNPs would indicate. We then identified these genomic regions in the Spring River hybrid samples that overlapped with the identified regions in the Tombigbee River hybrid samples to determine if patterns of introgression were the same across the two hybrid zones. Overlapping regions occurred in the same location on the same chromosomes for both river samples. For areas of low introgression, overlapping regions were identified on chromosomes 5, 10, 12, 15, and 21. For areas of high introgression, one overlapping region was identified and was located on chromosome 3. These regions were mapped to the *F. heteroclitus* annotated genome, which consists of scaffolds of varying lengths (eugene.org). Scaffolds over 10,000 base pairs in size were chosen for further study. Some scaffolds were completely within the regions, while others only had a portion inside of the region under investigation. If the scaffold began before the chosen region but did contain the region, the starting point of the scaffold was subtracted from the starting point of the chosen region to find how far into the scaffold to begin analysis on the genome browser. In other instances, if the chosen region ended before the scaffold ended, the starting point of the scaffold was subtracted from the ending point of the chosen region to find where to stop analysis on the genome browser. Within each scaffold, killifish genes with putative functional roles were noted. The genes' functions were identified using the Gene Ontology (GO) knowledgebase (uniprot.org). After analyzing the genes, it was determined that functions could be grouped into the following categories: cell cycle, cell development (differentiation), cell signaling, immune system, intra/intercellular transport, metabolism, transcription initiation/regulation, and translation initiation/regulation.

Chapter 3: Results

Structure Analysis

Datasets for individuals in both rivers were filtered by SNP and individual as described above. In addition, Structure assumes loci are not linked, so SNPs were thinned so that they were a minimum of 100 base pairs apart. In practice, SNPs are highly clustered and this level filtering results in larger distances among SNPs and satisfies linkage assumptions. In the Tombigbee River, genotyping 248 individuals resulted in 136,000 SNPs before filtering, and 192 individuals and 9,256 SNPs after filtering. In the Spring River, genotyping 111 individuals resulted in 296,000 SNPs before filtering, and 103 individuals and 9,744 SNPs after filtering. Structure analysis (K=2) of Tombigbee individuals identified 99 from parental species (46 individuals with admixture >0.99 and 53 with admixture <0.01). The remaining 93 individuals (48%) with intermediate admixture scores were considered to be of mixed ancestry (Fig. 1). In the Spring River, there were 70 individuals from parental species (36 with admixture >0.99 and 34 with admixture <0.01) and the remaining 33 (32%) were considered to be of mixed ancestry (Fig. 2). Only individuals from parental species were used in sliding window analyses described below.

Figure 1: Structure Analysis of the Tombigbee River Hybrid Zone Samples. The vertical bars represent an individual. Blue represents *F. olivaceus*, and red represents *F. notatus*. Individuals of mixed ancestry are seen in the middle where both blue and red are present on the bars.

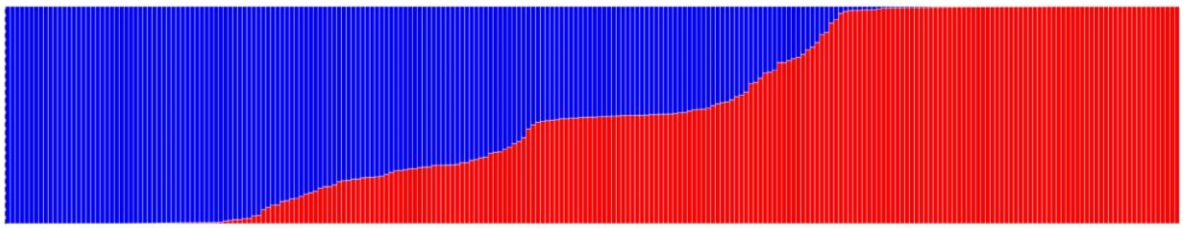
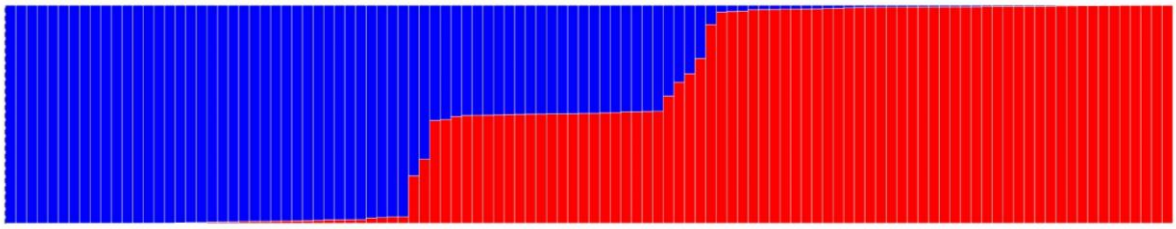


Figure 2: Structure Analysis of the Spring River Hybrid Zone Samples. The vertical bars represent an individual. Blue represents *F. olivaceus*, and red represents *F. notatus*. Individuals of mixed ancestry are seen in the middle where both blue and red are present on the bars.



Sliding Window Analysis

Because sliding window analyses do not assume SNPs in linkage disequilibrium, analyses started with the unfiltered SNPs for all individuals identified as parental in the two river systems. We used the program MUMmer to align the contigs (n=1,253) in the *F. olivaceus* reference (used in GBS genotyping samples with TASSEL) to the annotated reference for *F. heteroclitus* that is organized into 24 linkage groups (chromosomes). We discarded any alignments that were shorter than 2,000 base pairs or had a quality score <99% indicating lower rates of similarity within aligned regions. Of the 1,253 contigs, 128 (~1%) had no quality alignments to the *F. heteroclitus* reference. Of the 1,125 that had quality alignments, 643 had a large number (mean of 134) of alignments predominantly (>95%) to a single chromosome. These also tended to be the larger contigs in the dataset that captured most of the SNPs. We retained all SNPs on these contigs and assigned each SNP a chromosome number and position based on alignment data. For the samples from the two river systems, there were 117,875 (Tombigbee) and 259,253 (Spring) SNPs with position data after alignment. These SNPs were then filtered

in the same way as described above, except that we did not thin SNPs based on position. The final datasets used in the sliding window analyses contained 20,979 (Tombigbee) and 23,953 (Spring) filtered SNPs with position data. We used VCFtools to conduct a sliding window analysis along each chromosome. Windows were 500,000 base pairs with a 100,000 base step. The weighted mean F_{st} values between the parental species were 0.862 and 0.911 in the Spring and Tombigbee Rivers respectively. In the Tombigbee River, there were 7,169 windows, of which 651 had F_{st} values >0.99. In the Spring River there were 7,215 windows, of which 215 had F_{st} values >0.99.

Figure 3: Sliding Window Analysis of the Hybrid Genomes Found in the Tombigbee River. Red dots represent SNPs with F_{st} values >0.99. Green dots represent SNPs with F_{st} values <0.03. The x-axis represents the location on the chromosome in base pairs. The y-axis represents the chromosome, with each grey row as a different chromosome.

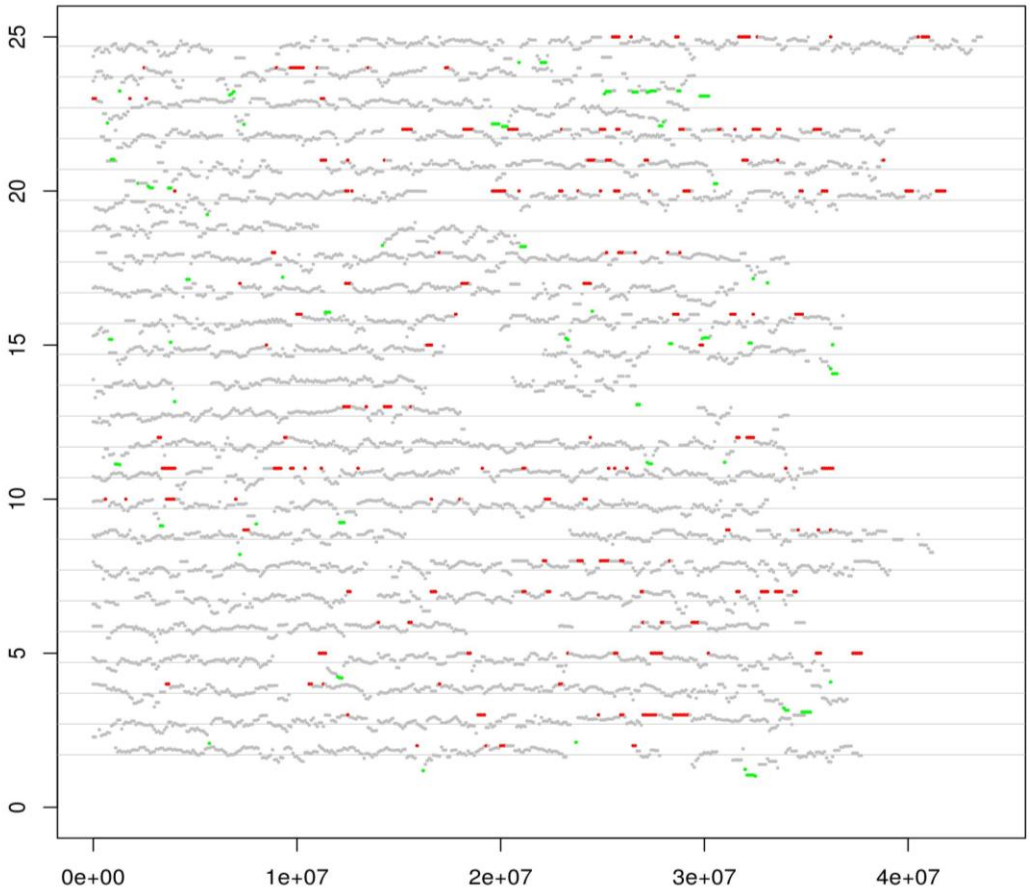
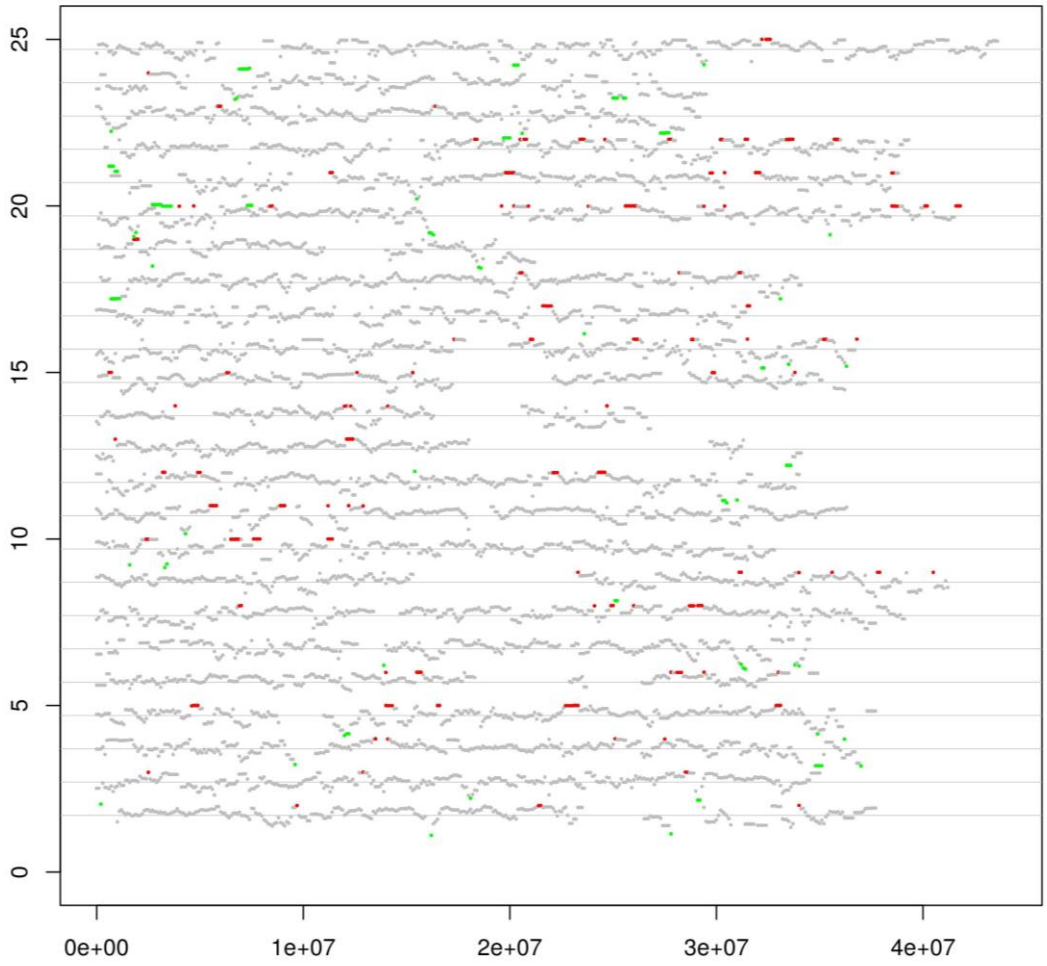


Figure 4: Sliding Window Analysis of the Hybrid Genomes Found in the Spring River. Red dots represent SNPs with F_{st} values >0.99. Green dots represent SNPs with F_{st} values <0.03. The x-axis represents the location on the chromosome in base pairs. The y-axis represents the chromosome, with each grey row as a different chromosome.



Hybrid Genome Analysis

To identify regions of low introgression, we looked for areas of the genome that had F_{st} values at or above 0.99 (Tombigbee) and 0.985 (Spring). Of these areas, we chose regions that had ten or more SNPs in two or more adjacent windows. These genomic regions that overlap between the Tombigbee and Spring river hybrid samples were then further analyzed. We identified regions of low introgression on 5 chromosomes, specifically chromosomes 5, 10, 12, 15, and 21. The location, number of

windows in the sliding window, and SNP numbers for each region are summarized in Table 1. Within these regions, there were a total of 97 genes with putative functions identified by using Gene Ontology (GO) knowledgebase. Their functions are summarized in Table 2. The largest number of genes were identified to have a function in cell development, followed by intra/intercellular transport and cell signaling.

Table 1: Locations of Overlapping Low Introgression Regions

Chromosome	**Region (base pairs)**	**# windows**	**SNPs (Tom.)**	**SNPs (Spring)**
5	27900000-28800000	5	36	16
10	8900000-9600000	3	9	12
12	12300000-12900000	2	12	9
15	10300000-10700000	2	9	14
15	26000000-26800000	4	23	24
15	29300000-29600000	4	9	20
15	35100000-35300000	2	9	4
21	20500000-21300000	4	11	15
21	35700000-36200000	1	6	8

Table 2: Gene Functions Identified in Regions of Low Introgression

Gene Function	**# Identified**
Cell Development	22 (22.7%)
Intra/Intercellular Transport	18 (18.6%)
Cell Signaling	17 (17.5%)
Cell Cycle	14 (14.4%)
Transcription Initiation/Regulation	13 (13.4%)
Metabolism	7 (7.2%)
Immune Function	4 (4.1%)
Translation Initiation/Regulation	2 (2.1%)

To identify regions of high introgression, we chose regions of the genome with F_{st} values at or below 0.03 that also had ten or more SNPs. From the overlapping regions

between the two hybrid zones, we identified regions of high introgression on only one chromosome, which was chromosome 3. There was one overlapping region identified on this chromosome, and it was found at 34800000-35600000 base pairs and contained 4 windows. This region contained 8 SNPs in the Tombigbee River hybrids and 16 SNPs in the Spring River hybrids. This region had 34 genes with known functions. Their functions are summarized in Table 3. The two categories with the largest number of genes identified were cell development and transcription initiation/regulation.

Table 3: Gene Functions Identified in Regions of High Introgression

Gene Function	**# Identified**
Cell Development	8 (23.5%)
Transcription Initiation/Regulation	8 (23.5%)
Cell Cycle	6 (17.6%)
Cell Signaling	4 (11.8%)
Intra/intercellular Transport	3 (8.8%)
Immune Function	3 (8.8%)
Metabolism	2 (5.9%)
Translation Initiation/Regulation	0 (0%)

Chapter 4: Discussion

In this study, we were interested in investigating the rates of hybridization and patterns of introgression in two hybrid zones found in the Tombigbee River and the Spring River. Determining the functions of genes located in portions of the genome exhibiting unusually high or low observed introgression was of particular interest. Genes in areas with low introgression are hypothesized to be important in maintaining species' boundaries. These areas of low introgression are considered genomic islands of speciation, as they are resistant to introgression and gene flow. Malinsky and his colleagues (2015) studied these islands of speciation in cichlid ecomorphs. These ecomorphs differ in body shape, male breeding color, diet, and depth preference. The researchers found nearly 100 genomic islands of speciation that contained adaptive genes involved in mate preferences, such as sensory perception, hormone signaling, and morphogenesis, which help maintain the species' boundaries present (Malinsky et al., 2015). The Spring River had a much greater percentage of parental species than species of mixed ancestry when compared to the Tombigbee River (Fig. 1). The Spring River had more F1 hybrids and fewer back crossed individuals, while the Tombigbee River had fewer F1 but many more back crosses. While the Spring River had more F1 hybrids, these seemed to have fewer offspring because there are fewer backcrosses. These results suggest that rates of hybridization may be similar in the two rivers, but F1 hybrids have a greater fitness in the Tombigbee River. This could be due to the difference in the geographical location and climate of the two hybrid zones. The Spring River is found at a higher elevation and experiences slightly cooler temperatures than the Tombigbee River. Another study was performed in the Glover and Cossatot Rivers containing *F.*

notatus and *F. olivaceus*, and similar levels of individuals with mixed ancestry were found between the hybrid zones studied (Schaefer et al., 2016). These two rivers are much closer to each other geographically and likely experience similar, if not the same, climates, which could possibly affect hybridization rates. The Tombigbee and Spring Rivers also differ in regards to the number of chromosomes each *Fundulus* species has. Typically, *F. notatus* has 40 chromosomes and *F. olivaceus* has 48 chromosomes (Duvernell et al., 2007), and we see this pattern in the Spring River. However, in the Tombigbee River, the *F. notatus* species has 44 chromosomes (Duvernell et al., 2007; Black & Howell, 1978). The *Fundulus* species present in the Tombigbee River are a unique chromosomal race (Black & Howell, 1978), perhaps due to fewer chromosomal fusions in its population of *F. notatus* compared to the population in the Spring River. Thus, the karyotypes of each species are more similar to each other in the Tombigbee River than in the Spring River. Less chromosomal differences in the Tombigbee River could explain the greater hybrid fitness seen.

With respect to patterns of introgression in the Tombigbee River and Spring River *Fundulus* hybrid zones, we found overall more genomic regions exhibiting low introgression versus high introgression. Specifically, we identified nine genomic regions showing low introgression that were present in both rivers, and these occurred within five chromosomes. These results are similar to a study of hybrid zones of *Melocactus* species (Khan et al., 2019). While the various species are likely to hybridize, genomic introgression rates were very low (Khan et al., 2019). There were similar low rates of high introgression between the two rivers, but only one region was identified that was in the same chromosomal location for both hybrid zones. This region was located on

chromosome 3. This suggests that genes in this region of chromosome 3 may have an adaptive value for both species in their environments. If a gene increases the fitness of an organism, that gene is likely to be passed to offspring. Since high introgression is present in this region, the genes from both species are being passed on, suggesting that they are beneficial. The *Fundulus* hybrid zones are very spatially limited, and there is very little evidence of hybrids a few kilometers away from the contact zone. This could be seen as strong selection against hybridization and less introgression compared to the cichlid system (Malinsky et al., 2015).

We also studied the functional roles of the genes found in the areas of high and low introgression identified. Hybridization and introgression can be very useful evolutionary tools as they increase genetic variability (Stelkens et al., 2014). Exchanging genes with related species can bring about genetic variation faster than the typical rate that mutations accumulate, thus the populations can adapt more quickly (Stelkens et al., 2014). In a study of two other *Fundulus* species, introgression was found to be very beneficial and brought about adaptive variation when the environment was very polluted (Oziolor et al., 2019). For areas of the genome exhibiting high introgression, we identified 34 functional genes within the one region. The two largest categories were cell development (8) and transcription initiation/regulation (8).

For low introgression, we identified 97 functional genes, with most falling into the category of cell development (22), followed by intra/intercellular transport (18) and cell signaling (17). Cell development could include, for example, genes that code for proteins that guide the coloration/pattern present on the fish, which is a key physical difference between the two species. The spots present on *F. olivaceus* are sexually

dimorphic, and *F. olivaceus* females prefer males with more spots, as the spots are an indicator of fitness (Schaefer et al., 2012). This is an honest signal from the males, and males with more spots are more prone to predation. It is also likely a visual cue because populations in more turbid water have fewer spots (Schaefer et al., 2012). Transcription and cell cycle genes could also contribute to these spots being sexually dimorphic and an honest signal. While we could have also expected to see low introgression in areas of the genome containing these genes, this was not the case. In a study of mule deer and black-tailed deer hybrid zones, the hybrids were found to have selection for genes associated with general cell processes, which relate to these three categories noted. They hypothesized that this selection may be due to environmental adaptation, as the hybrid zone occurred in an area with a sharp ecological transition (Haines et al., 2019). In another study regarding field cricket hybrid zones, genes with restricted introgression were mostly encoding cytoskeletal proteins and growth-stimulating proteins (Larson et al., 2013), which were both considered to belong to the cell development category for our study. Larson further speculated that cytoskeletal proteins are often involved in fertilization and gamete formation processes, and the growth-stimulating proteins could have been involved in differences in body size of the species under study (Larson et al., 2013). Some cell development genes identified in these regions of low introgression could have a role in gamete production and reproduction, which are different in *F. olivaceus* and *F. notatus.* *Fundulus olivaceus* is also slightly larger, and cell development genes could potentially have a role in this difference as well.

We somewhat expected to see low introgression in regions that contain genes associated with immune function. This expected pattern of low introgression in immune

function genes could have been associated with the fact that these two species occupy different habitats and come in contact with different types and amounts of pathogens. Previous studies in salmonids have indeed demonstrated that different populations of the same species were subject to differential pathogen pressures, resulting in allele frequency differences at the major histocompatibility complex (MHC), which plays an integral role in pathogen recognition (Evans et al., 2010). However, we did not see this trend between these two specific hybrid zones. In the same deer hybrid zones study, Haines and her colleagues (2019) also expected that immune function genes would be under selection, with different alleles being selected for in the two species. A large number of immune system genes have been seen to be under selection in various mammals (Roffler et al., 2016; Gouveia et al., 2017; Gautier et al., 2009; Schweizer et al., 2016). However, their study showed that there was only one SNP likely to be under this selection (Haines et al., 2019). This is similar to our results, where we only saw three genes with immune functions in the notated regions of low introgression.

Further studies could compare these hybrid zones to more closely positioned hybrid zones in other rivers. This could give more insight as to whether geographical or environmental factors influenced our results. It would also be interesting to perform a further analysis of the specific genes we identified and analyze differentiation at each loci. Certain specific genes could be present in regions of low introgression that could be considered important for maintaining species' boundaries.

References

Abbott, R., Albach, D., Ansell, S., Arntzen J.W., Baird, S.J.E., Bierne, N., Boughman, J., Brelsford, A., Buerkle, C.A., Buggs, R., Butlin, R.K., Dieckmann, U., Eroukhmanoff, F., Grill, A., Cahan, S.H., Hermansen, J.S., Hewitt, G., Hudson, A.G., Jiggins, … Zinner, D. (2013). Hybridization and speciation. *Journal of Evolutionary Biology*, *26*(2), 229–246. doi:10.1111/j.1420-9101.2012.02599.x

Abbott, R.J., Barton, N.H. & Good, J.M. (2016). Genomics of hybridization and its evolutionary consequences. *Molecular Ecology*, *25*(11), 2325-2332. doi:10.1111/mec.13685

Arnold, M. L. (1992). Natural hybridization as an evolutionary process. *Annual Review of Ecology and Systematics*, *23*(1), 237–261. https://doi.org/10.1146/annurev.es.23.110192.001321

Black, A., & Howell, W.M. (1978). A distinctive chromosomal race of the Cyprinodontid fish, *Fundulus notatus*, from the upper Tombigbee River system of Alabama and Mississippi. *Copeia, 1978*(2), 280–288. DOI: 10.2307/1443564

Brennan, R. S., Healy, T.M., Bryant, H.J., La, M.V., Schulte, P.M., & Whitehead, A. (2018). Genome-wide selection scans integrated with association mapping reveal mechanisms of physiological adaptation across a salinity gradient in killifish. bioRxiv 254854.

Dittrich-Reed, D. R., & Fitzpatrick, B.M. (2013). Transgressive hybrids as hopeful monsters. *Evolutionary Biology 40*(2), 310–315. https://doi.org/10.1007/s11692-012-9209-0

Duvernell, D.D., Meier, S.L., Schaefer, J.F., & Kreiser, B.R. (2013). Contrasting phylogeographic histories between broadly sympatric topminnows in the *Fundulus notatus* species complex. *Molecular Phylogenetics and Evolution, 69*(3), 653–663. 10.1016/j.ympev.2013.07.013

Duvernell, D. D., & Schaefer, J. F. (2013). Variation in contact zone dynamics between two species of topminnows, *Fundulus notatus* and *F. olivaceus*, across isolated drainage systems. *Evolutionary Ecology 28*, 37–53. 10.1007/s10682-013-9653-z

Duvernell, D.D., Schaefer, J.F., Hancks, D.C., Fonoti, J.A. & Ravanelli, A.M. (2007). Hybridization and reproductive isolation among syntopic populations of the topminnows *Fundulus notatus* and *F. olivaceus*. *Journal of Evolutionary Biology, 20*(1), 152-164. doi:10.1111/j.1420-9101.2006.01213.x

Elshire, R. J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., and Mitchell, S.E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE, 6*(5), e19379. https://doi.org/10.1371/journal.pone.0019379

Evans, M., Neff, B. & Heath, D. (2010). MHC genetic structure and divergence across populations of Chinook salmon (*Oncorhynchus tshawytscha*). *Heredity, 104*(5)**,** 449–459. https://doi.org/10.1038/hdy.2009.121

Gautier, M., Flori, L., Riebler, A., Jaffrezic, F., Laloe, D., Gut, I., Moazami-Goudarzi, K., & Foulley, J.L. (2009). A whole genome Bayesian scan for adaptive genetic divergence in West African cattle. *BMC Genomics, 10*(550). https://doi.org/10.1186/1471-2164-10-550

Glaubitz, J.C., Casstevens, T.M., Lu, F., Harriman, J., Elshire, R.J., Sun, Q., & Buckler, E.S. (2014). TASSEL-GBS: A high capacity genotyping by sequencing analysis pipeline. *PLoS ONE, 9*(2), e90346. https://doi.org/10.1371/journal.pone.0090346

Gouveia, J.J.D., Paiva, S.R., McManus, C.M., Caetano, A.R., Kijas, J.W., Facó, O., Azevedo, H.C., de Araujo, A.M., de Souza, C.J.H., Yamagishi, M.E.B., Carneiro, P.L.S., Lôbo, R.N.B., de Oliveira, S.M.P., & da Silva, M.V. (2017). Genome-wide search for signatures of selection in three major Brazilian locally adapted sheep breeds. *Livestock Science, 197,* 36–45. https://doi.org/10.1016/j.livsci.2017.01.006

Haines, M.L., Luikart, G., Amish, S.J., Smith, S., & Latch, E.K. (2019). Evidence for adaptive introgression of exons across a hybrid swarm in deer. *BMC Evolutionary Biology, 19***(**199). https://doi.org/10.1186/s12862-019-1497-x

Khan, G., Franco, F.F., Silva, G.A.R., Bombonato, J.R., Machado, M., Alonso, D.P., Ribolla, P.E.M., Albach, D.C., & Moraes, E.M. (2019). Maintaining genetic integrity with high promiscuity: Frequent hybridization with low introgression in multiple hybrid zones of Melocactus (Cactaceae). *Molecular Phylogenetics and Evolution*, *142*, 106642. 10.1016/j.ympev.2019.106642.

Langmead, B., & Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, *9*, 357–359. https://doi.org/10.1038/nmeth.1923

Larson, E.L., Andrés, J.A., Bogdanowicz, S.M. & Harrison, R.G. (2013), Differential introgression in a mosaic hybrid zone reveals candidate barrier genes. *Evolution, 67*(12), 3653-3661. doi:10.1111/evo.12205

Mallet, J. (2005). Hybridization as an invasion of the genome. *Trends in Ecology & Evolution, 20*(5), 229-237. DOI: 10.1016/j.tree.2005.02.010.

Malinsky, M., Challis, R.J., Tyers, A.M., Schiffels, S., Terau, Y., Ngatunga, B.P., Miska, E.A., Durbin, R., Genner, M.J., & Turner, G.F. (2015). Genomic islands of speciation separate cichlid ecomorphs in an East African crater lake. *Science, 350*(6267), 1493-1498. DOI:10.1126/science.aac9927

Marçais, G., Delcher, A.L., Phillippy, A.M., Coston, R., Salzberg, S.L., and Zimin, A. (2018). MUMmer4: A fast and versatile genome alignment system. *PLOS Computational Biology, 14*(1), e1005944. doi: 10.1371/journal.pcbi.1005944

Oziolor, E.M., Reid, N.M., Yair, S., Lee, K.M., VerPloeg, G.S., Bruns, P.C., Shaw, J.R., Whitehead, A., & Matson, C.W. (2019). Adaptive introgression enables evolutionary rescue from extreme environmental pollution. *Science, 364*(6439), 455-457. DOI: 10.1126/science.aav4155

Pritchard, J.K., Stephens, M., & Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics, 155*(2), 945–959.

Roffler, G.H., Amish, S.J., Smith, S., Cosart, T., Kardos, M., Schwartz, M.K., & Luikart, G. (2016). SNP discovery in candidate adaptive genes using exon capture in a freeranging alpine ungulate. *Molecular Ecology Resources, 16*(5), 1147–1164. doi: 10.1111/1755-0998.12560

Schaefer, J., Duvernell, D., & Campbell, D.C. (2016). Hybridization and introgression in two ecologically dissimilar *Fundulus* hybrid zones. *Evolution, 70*(5), 1051–1063. doi: 10.1111/evo.12920

Schaefer, J.F., Duvernell, D.D., & Kreiser, B.R. (2011). Ecological and genetic assessment of spatial structure among replicate contact zones between two topminnow species. *Evolutionary Ecology, 25*, 1145–1161. 10.1007/s10682-011-9461-2

Schaefer, J. F., Duvernell, D.D., Kreiser, B.R., Champagne, C., Clark, S.R., Gutierrez, M., Stewart, L.K., & Coleman, C. (2012). Evolution of a sexually dimorphic trait in a broadly distributed topminnow (*Fundulus olivaceus*). *Ecology and Evolution, 2*(7), 1371–1381. https://doi.org/10.1002/ece3.242

Schweizer, R.M., Robinson, J., Harrigan, R., Silva, P., Galverni, M., Musiani, M., Green, R.E., Novembre, J., & Wayne, R.K. (2016). Targeted capture and resequencing of 1040 genes reveal environmentally driven functional variation in grey wolves. *Molecular Ecology, 25*(1), 357–379. https://doi.org/10.1111/mec.13467

Seutin, G., White, B.N., & Boag, P.T. (1991). Preservation of avian blood and tissue samples for DNA analyses. *Canadian Journal of Zoology, 69*(1), 82–90. DOI: 10.1139/z91-013

Stelkens, R.B., Brockhurst, M.A., Hurst, G.D.D. & Greig, D. (2014). Hybridization facilitates evolutionary rescue. *Evolutionary Applications, 7*(10), 1209-1217. doi:10.1111/eva.12214

Twyford, A.D., & Ennos, R.A. (2012). Next-generation hybridization and introgression. *Heredity*, *108*(3), 179–189 https://doi.org/10.1038/hdy.2011.68

Waits, E., Martinson, J., Rinner, B., Morris, S., Proestou, D., Champlin, D. & Nacci, D. (2016). Genetic Linkage Map and Comparative Genome Analysis for the Atlantic

Killifish (*Fundulus heteroclitus*). *Open Journal of Genetics, 6*(1), 28-38. doi: 10.4236/ojgen.2016.61004.

Vernot, B., & Akey, J.M. (2014). Resurrecting surviving neandertal lineages from modern human genomes. *Science, 343*(6174), 1017–1021. 10.1126/science.1245938